

Cognome e Nome _____

Matricola _____

Big data – Graph Mining A.A. 2025/2026

Appello del 15 luglio 2026

Esercizio 1 [8 punti]

Sia dato un grafo diretto di transazioni monetarie, pesato sugli archi. Un nodo corrisponde ad una persona e il suo peso al suo saldo iniziale. Un arco diretto dal nodo i al nodo j avente peso w indica una transazione che comporta trasferimento di w denaro dalla persona i alla persona j . Le transazioni, tuttavia, sono tassate, e quando si trasferisce w denaro dalla persona i alla persona j il saldo della persona i diminuisce di w , mentre il saldo della persona j aumenta del **90% di w** (il restante 10% corrisponde alla tassazione). Tutte le persone hanno, inizialmente, un saldo pari a **1000**.

Si progettino (eventualmente utilizzando più step di computazione tra loro concatenati) le funzioni *Map* e *Reduce* per restituire in output **tutte le persone aventi come saldo finale il massimo saldo finale tra i saldi finali di tutte le persone del grafo** dato un **grafo diretto pesato su nodi e archi**, utilizzando spazio locale $O(n^{1/2})$, dove n è il numero di nodi del grafo. Si assuma che i nodi siano rappresentati da numeri interi da 0 a $n-1$ e che un arco del grafo, che va dal nodo i al nodo j ed ha peso w , sia codificato dalla tripla (i, j, w) . Si assuma anche che ogni nodo sia coinvolto in un numero costante di transazioni, e che ogni nodo riceva almeno una transazione in ingresso.

Suggerimento: si utilizzi un primo step di computazione per calcolare il saldo finale di ogni nodo; successivamente, si utilizzino gli step di computazione necessari a trovare il massimo saldo finale.

Esempio di input:

- (0, 1, 300)
- (2, 1, 200)
- (0, 2, 100)
- (2, 0, 100)

- (0, 690)
- (1, 1450)
- (2, 790)

Per poi calcolare tutte le persone associate al massimo saldo finale (solo 1 nell'esempio, che è associata al saldo 1450).

In questo input abbiamo 3 nodi e 4 archi.

Il suggerimento consiglia di calcolare prima le triple

(id_nodo, saldo_finale)

che nell'esempio risulteranno essere:

Esercizio alternativo [valutazione massima 6 punti]: svolgere l'esercizio con l'unica modifica che bisogna dare in output **una qualsiasi persona avente come saldo finale il saldo finale massimo** (sempre utilizzando spazio locale $O(n^{1/2})$).

Esercizio alternativo [valutazione massima 4 punti]: assumere di avere in input le coppie **(id_nodo, saldo_finale)** (con id_nodo che varia da 0 a $n-1$) e dare in output **una qualsiasi persona avente come saldo finale il saldo finale massimo, utilizzando spazio locale $O(n^{1/2})$** .

Esercizio 2 [9 punti]

Si consideri il problema dei **Frequent Itemset**.

- a. Si definisca il **bordo negativo** e si indichi come è possibile calcolarlo
- b. Si descriva il funzionamento dell'**algoritmo di Tovoinen**
- c. Si definiscano i **falsi negativi** e i **falsi positivi**, e si discuta come vengono gestiti dall'algoritmo di Tovoinen

Esercizio 3 [9 punti]

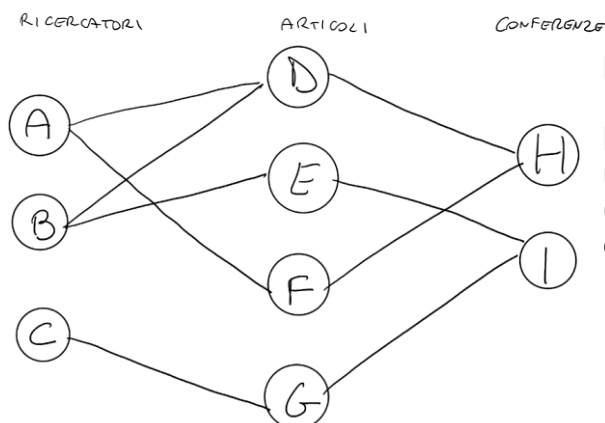
Si considerino gli insiemi descritti nella tabella qui sotto.

- **[3 punti]** Si scriva la colonna dell'insieme S_4 in modo che la Jaccard Similarity tra S_3 e S_4 sia $1/6$. Si calcoli quindi la Jaccard Similarity tra S_1 e S_4 e tra S_2 e S_4 .
- **[4 punti]** Si calcoli, applicando il metodo visto a lezione ed evidenziando tutti i passaggi fatti riga per riga, la matrice delle firme **minhash** degli insiemi S_1 , S_2 , S_3 e S_4 secondo le permutazioni indotte dalle 3 funzioni hash evidenziate in tabella.
- **[2 punti]** Sfruttando la matrice delle firme appena calcolata, si approssimi la Jaccard Similarity tra S_1 e S_4 , tra S_2 e S_4 e tra S_3 e S_4 .

Riga	S_1	S_2	S_3	S_4	$(3x+1) \bmod 7$	$(x+3) \bmod 7$	$(2x-2) \bmod 7$
0	1	0	1				
1	0	1	1				
2	0	0	0				
3	1	1	0				
4	0	0	1				
5	0	1	0				
6	1	1	1				

Esercizio 4 [8 punti]

Si consideri la rete sociale in figura, che rappresenta con un grafo 3-partito le relazioni tra ricercatori, articoli scientifici e conferenze: l'insieme di nodi più a sinistra rappresenta i ricercatori, quello centrale gli articoli e quello di destra le conferenze; un arco tra un ricercatore e un articolo indica il fatto che il ricercatore è autore dell'articolo, mentre un arco tra un articolo e una conferenza indica il fatto che l'articolo è stato presentato alla conferenza.



[2 punti] Come è definito, nel SimRank, l'insieme di teleport?

[6 punti] Si impostino le equazioni per il calcolo del **SimRank** rispetto al ricercatore **B** con parametro di teleport β generico (si noti che il grafo è non diretto, e quindi ogni arco va considerato come doppiamente diretto).

Regole per lo svolgimento della prova scritta:

- Per svolgere il compito si hanno a disposizione **90 minuti (70 minuti se non si svolge l'Esercizio 1)**
- Scrivere **subito** nome, cognome, matricola su **OGNI FOGLIO (compreso questo)**.
- Durante la prova scritta **non** è possibile abbandonare l'aula.
- Non è ammesso **per nessun motivo** comunicare in qualsiasi modo con altre persone
- Non è possibile consultare appunti, libri e dispense.
- Qualsiasi strumento elettronico di calcolo o comunicazione (telefoni cellulari, calcolatrici, tablet, computer, smartwatch, etc...) deve essere **completamente disattivato** e **depositato in vista sulla cattedra**
- Mettere in vista sul banco un valido documento di identità.